

## Appendix 1. BLAST alignments

### A. hepsin vs spinesin:

NCBI Blast 2 Sequences results

PubMed Entrez BLAST OMIM Taxonomy Structure

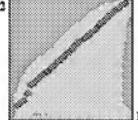
BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.17 [Aug-26-2007]

Matrix [BLOSUM62] gap open [1], gap extension: [1]  
X-dropoff: [0] expect: [0.000] wordsize: [3] Filter: [C] View option: [Standard]  
Masking character option: [X for protein, n for nucleotide] Masking color option: [Black]  
Show CDS translation: [Align]

---

Sequence 1: gi|2120578|sp|Q00917|hepsin protease hepsin [Trinucleotide protease, serine 1] [Contains: Serine protease hepsin non-catalytic chain; Serine protease hepsin catalytic chain]  
Length = 417 (1 .. 417)

Sequence 2: gi|17240917|sp|Q9H8W1|spinesin [Homo sapiens]  
Length = 457 (1 .. 457)



NOTE: Bitscore and expect value are calculated based on the size of the nr database.

---

Score = 211 bits (563), Expect = 1e-55  
Identities = 140/412 (34%). Positives = 200/412 (48%). Gaps = 81/412 (19%).

Query 24 G T L I L I T A G A A N V A V A V A T L S R G Q E P L Y P V P ----- 69  
Subject 63 G A / C L I G A Q J A S T S W L V I L C G R A H Q R Q S T T L Q O Z E X T I L S C F I S R A R A L L P A L M T V S S 112

Query 70 ----- 67 G T W P L L C S S R D R N A R Y T A G L C E R E M F D P A L T H T E L D V T R A R A N G T 114  
Subject 118 R I N S E K F I L A R E N P D O S P R P L U V C H E G N E R A I S Q U T H S D A L A L T H E R G N I T D I K L N S S 172

Query 118 R E P C U T E B R G A L P H T Q G A L I V I U -- S V C D P P R P T L A I C Q C G C K A Z L P V D P R V Q Q D R T S L S 171  
Subject 173 Q S F A C I S ----- F R Q G P Z L E Z A M D P P M N C T S Q W V I R C U C E T T A P P L - A S P R Y G L Q S V V A P G 217

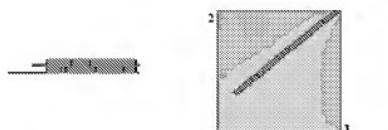
Query 173 R M V Q V P L R X D R A M L Q Q 9 7 S L S G U W I T R A N C P F E - R M V L S R M G Y F A G A V V Q A S - E H S 229  
Subject 229 A M P K Q S + H C Q G S A - L R W V T A R T R L S W V V A G V + ++ R H



## B. hepsin vs TMPRSS2:

**Sequence 1** gi:323037|Serine protease heparin (Transmembrane protease, serine 1) [Contains Serine protease heparin non-catalytic chain, Serine protease heparin catalytic chain]  
Length = 417 (1 .. 417)

Sequence 1: gr14607456-transmembrane protease, serine 2 [Homo sapiens]  
gr30017246-peptidase, peptidase M23, transmembrane serine protease 2 [Homo sapiens]  
Length = 492 (I..492)

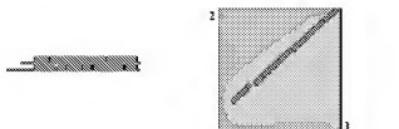


**NOTE:** Hitcount and exact value are calculated based on the size of the database.

### C. hepsin vs TMPRSS3:

**Sequence 1:** g1:1-3057|Serine protease hepsin (Transmembrane protease, serine 3) [Contains: Serine protease hepsin non-catalytic chain; Serine protease hepsin catalytic chain]  
Length = 417 (1 .. 417)

Sequence 2: g1371821821MPRSS3 [Homo sapiens]  
Length = 432 (1..432)



**NOTE:** Bitacose and expect value are calculated based on the size of the m1 database.

## D. hepsin vs TMPRSS4:

Sequence 1. gi|13807|Serine protease hepsin (Transmembrane protease, serine 11 [Contains: Serine protease hepsin non-catalytic chain; Serine protease hepsin catalytic chain])  
Length = 417 (1 .. 417)

Sequence 2. gi|745194|transmembrane protease, serine 4 isoform 1 [Homo sapiens]  
\_gi|12653890|sp|Q07847|TPS4\_HUMAN Transmembrane protease, serine 4 (Membrane-type serine protease 2) (MTP-SPase); gi|832745|gb|AAF74026|TSP24\_1 transmembrane serine protease 3 [Homo sapiens]; gi|5019793|gb|AAH17704|Transmembrane protease, serine 4 [Homo sapiens]; gi|2077743|gb|BAD18749|1 unnamed protein product [Homo sapiens].  
Length = 437 (1 .. 437)



NOTE Bitscore and expect value are calculated based on the size of the nr database.



Score = 213 bits (558), Expect = 4e-55  
Identities = 126/387 (32%), Positives = 180/387 (46%), Gaps = 23/387 (6%)  
Query 90 PLYPVWSEADARMVPTPTEGTWNLLCGSERHARAVLGLCDEEDMFLRALTSHEDVITA 109  
P V+S + L V D G N C +K+ H+H+ T +++  
Subject 166 PAVAVDLEKEDRDTQLVLDLRRNNSPACYDNSTERLAKTACRNQHISSEKPTPRANEI--- 186  
Query 110 RANTTSQTTTCVVEGRPHPTTQHMLAEVYITVTDQVWVQVTLAALICDQDSDPRLVTRVNGRDI 195  
G + + + E + +P+L + C G + + C D+ I + +H+G+  
Subject 187 QGDQDLDIVETIZE----HEDZLMRNNSSQGCLQEGIYVNLCLAOQH-EHTTPRNGKEA 211  
Query 176 SLCGNMPWQVFLYDQJAHLSLQSLCENQVLTAAACSYTPEVNNVLEDRNNVPGAGVAQCRSPW 228  
S+ NDNQW/64+2D H+DGS+L MVLTRANCE + U + H+V AQ+ S  
Subject 212 SVDENRQVQEITQYDQVMQWQGQSIIDPNVVLYRANCTRHIDVPUH-MEVRAASIMNGSTPS 270  
Query 220 LQLOVQDNYVHPQHILPFRDNESEKNDIQLWHESSFSLHIIIPQVULPRAA-QRIVYDK 289  
L + + + + P HCDIA L P+ + +P+CILF + +  
Subject 271 LAVKVKI2IIFNPMH-----KNDIAKHKQFELITGIVVRSICLQFFPFEELPAT 322  
Query 230 IOTVIGWMT-UNVPPQGAGWVQADWVIIKNDVNCAGAFGMQGHMSNPVAYPREGDVA 348  
+ GNG I Q G+ +L +K V +I + C H D Y ++ PM CRG PEG+D  
Subject 312 PLHIONQDFTYVQHGQHNRDILLQAVQVIESTRNADCAVQGEVTERDQHACIPLGGDVT 381  
Query 249 QDIDSPDIPFVCE0618RTRNLKGQZIWEWQTCQCALAQNSPVVTKNSDFPENIQPAK 465  
CQGDDGDF + + + +H + GIVNG GC PGIYIKS + RIV Y  
Subject 388 CQGDDGDFVLNHC-----SDQWRYHPIVWNGVQGQGQGCTPQVITVQGAIENNNINNG 434

E. hepsin vs enteropeptidase:

**Sequence 1.** [gi|276077](#)Serine protease hepsin (Transmembrane protease, serine 1) [Contains Serine protease hepsin non-catalytic chain, Serine protease hepsin catalytic chain]  
Length = 417 (1 .. 417)

**Sequence 2.** [gi|600901](#)enteropeptidase [Homo sapiens]  
Length = 1619 (1 .. 1619)



NOTE Bitscore and expect value are calculated based on the size of the nr database.

```
Score = 111 bits (53%), Expect = 1e-52
Identities = 108/256 (42%), Positives = 145/256 (56%), Gaps = 17/256 (6%)
Query 151 QDCGRKRLWD---PIVQXPQDTIAGRNIVNQVQEVLRDKDAMLCGQ972SLSGIVNTLTANCFVY 207
+ CG++ D +IVQXPQDTIAGRNIVNQVQEVLRDKDAMLCGQ972SLSGIVNTLTANCFVY 207
Subject 770 KSCGGKGLAAQCDIKRIVGVSNNMEGQANPRVTVGKLYXQSPRLQASILSVSMEWVSEARFCVY 829

Query 202 PNPVLEPNMTAGAYVAGA---EPWGLOLPVQADPVTWGV197FDQWNEEERKNDALVWLS 264
RS -3W G ++ EF - - - - 4V + Y B HDEA--WLS 264
Subject 935 PNLEP83NMTAIIGLHMEENLNLTPQTVPALIDEKVINPHY-----NSPRAVNDCLIAQGHL 385

Query 268 SFLSLTEYDOPVLPRAQQLAVVQGKICIVTWNQNTVYYQQQAVLQJQARVBLIIRHNVVNG 324
+ T-YDOP-CLP Q G C+ 2W G R -1QSA VP+-H+H C 324
Subject 934 FIKQVUDIVDQICLPEHQVQFFPGRHR781XNWGTUVQQTATIHLQRADTELLSNERCQ- 942

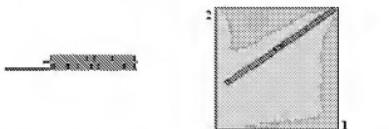
Query 325 ADFYQNGQIKFPRHFCAGYEGGIDACQZD8GOFWVCEI618RIPFWRLOGI78NOTVCALA 384
E M QMY E2G1DHQOQEGSF +D+-- SW L G+ 84G CAL

Subject 943 QQMPEINITENMICRAGYEGGIDACQZD8GOFWVCEI618RIPFWRLOGI78NOTVCALA 384
Dmtry 335 QKPGQYTKVBDPPEWIK 400
+PGQY +VE F SWI
Subject 955 WAGVTAARNSEPFIZI 1014
```

#### F. hepsin vs MSPL:

Sequence 1: gi|223057|Serine protease hepsin (Trans-membrane protease, serine 23) [Contains: Serine protease hepsin non-catalytic chain; Serine protease hepsin catalytic chain]  
Length = 417 (1 .. 417)

Sequence 2: g1116256163;transmembrane protease, renin 13 [Homo sapiens]  
Length = 567 (1 .. 567)



NOTE Bitscore and expect value are calculated based on the size of the nr database.

www.ijerph.org

Score = -151 bits (588), Expect = 1e-53  
Identities = 124/372 (33%), Positives = 196/372 (52%), Gaps = 19/372 (5%)

Query 55 LTHSELDVTRAGANTSGEFTVDEKRLPNTORLLEVIEVSOCBEPGRALAZICQDGCRRLX 168  
-P4 P 7 L 4 E + ACB Beta C CG R +

Quality 1.000 EVERLIVINGDPTSLGRMPNQCVSLEIDGAHL202SSLLSGCNYLTAACG-EMBRBNEVLSRNVY

Subject: 323 - TERRIGALLASDEKWTWQVSLHFTTICOGTLLDAQVLTAAHTFFVRENVLGCKV 39 L

+AG + + H L + + + + ? ZZ+ DIAL+ LS PL L+ + I P

Project 882 YRGTC--SULSQLPEASIAEILINSNY----TGEZRDYDIALMRSLXPKTLSAHMP 482  
Oncore 276 INCLBLAFCOMIHDCKLTGTDWQHONWQHODG-AHDFRBDLLEHTMNDLSDWVWQHKA 324

CLP GO + C +D94G D+ + L4E 4V +E CN Y + P  
Object 483 ALCMHNQQTISLNENNTWIDGPFYRTEDDKTSPTLREVQWNLIDFHVNDYLYNDVSLTP 481

Query 886 NTMFGYEGGIDAQDQDPGGPPEQEDLIRRTPHMLRIGVNGICLQLQKQGIVRVS  
+M DAPG GC D-CQGDSQGE VCE + PE L Q+ SNGTGC KRGWVTRV+  
S86 886 PEGGIDAQDQDPGGPPEQEDLIRRTPHMLRIGVNGICLQLQKQGIVRVS  
P86 886 PEGGIDAQDQDPGGPPEQEDLIRRTPHMLRIGVNGICLQLQKQGIVRVS

Query 166 DFREWITQAIKI 466  
+ KI+ ++